Center for
Basic **A**lgorithms **R**esearch **C**openhagen

2017 VILLUM Investigator

Mikkel Thorup

Information and Communication Technologies (ICT)

- a vehicle for societal progress

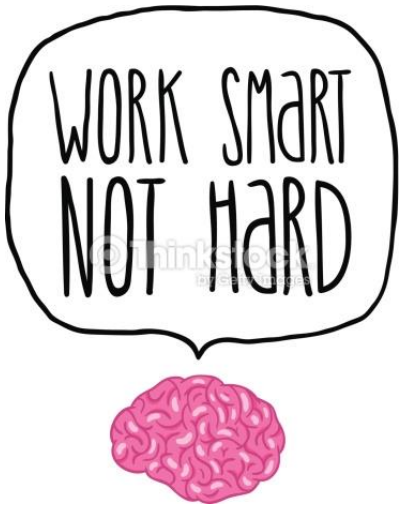
"Information and Communication Technologies underpin innovation and competitiveness across private and public sectors and enable scientific progress in all disciplines."

"The ICT sector represents 4.8% of the European economy."

"Investments in ICT account for 50% of all European productivity growth."

Algorithms are key to keeping up progress

The amount of data grows much faster than computer speeds, so need for efficient algorithms to process data becomes more and more urgent.



Example: Vimeo

Main competitor of YouTube – 170 million users/month.
Serves about 1 billion requests for video clips per day.

Key technology: Consistent hashing

Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications

Ion Stoica*, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan†
MIT Laboratory for Computer Science
chord@lcs.mit.edu
http://pdos.lcs.mit.edu/chord/

Abstract

A fundamental problem that confronts peer-to-peer applications is to e
pap
this
a k
imp
iter
key
sys
cha
per
and the state maintained by each node scaling logarithmically with the number of Chord nodes.

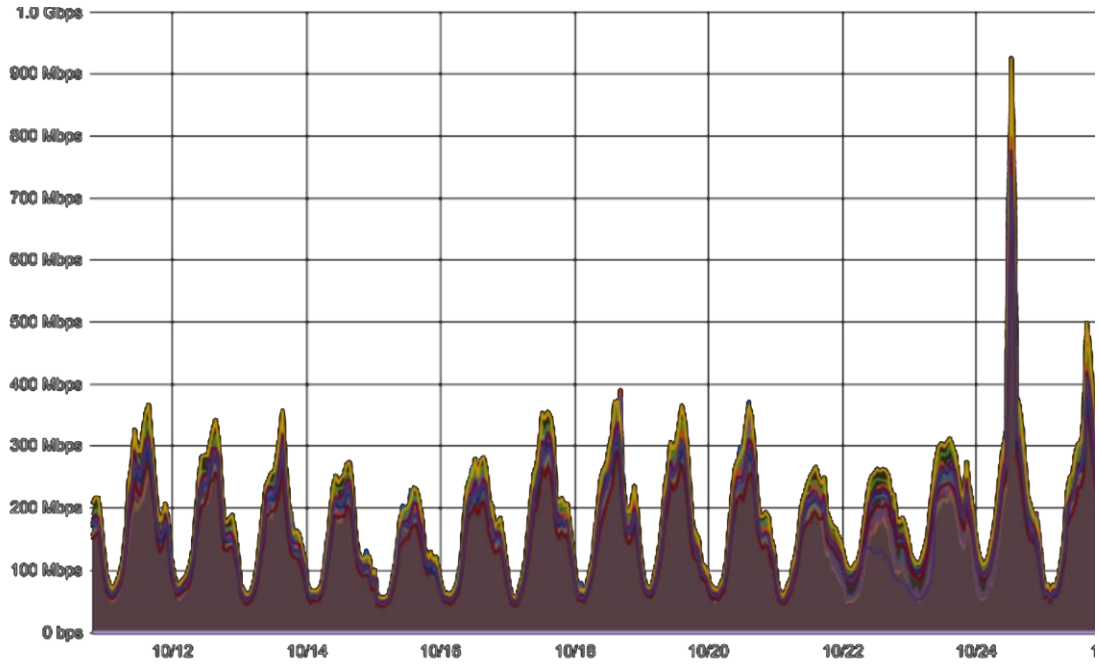
and involves relatively little movement of keys when nodes join and leave the system.

Title 1–20	Cited by	Year
Chord: A scalable peer-to-peer lookup service for internet applications I Stoica, R Morris, D Karger, MF Kaashoek, H Balakrishnan ACM SIGCOMM Computer Communication Review 31 (4), 149-160	12552	2001

event results in no more than $O(\log^2 N)$ messages.

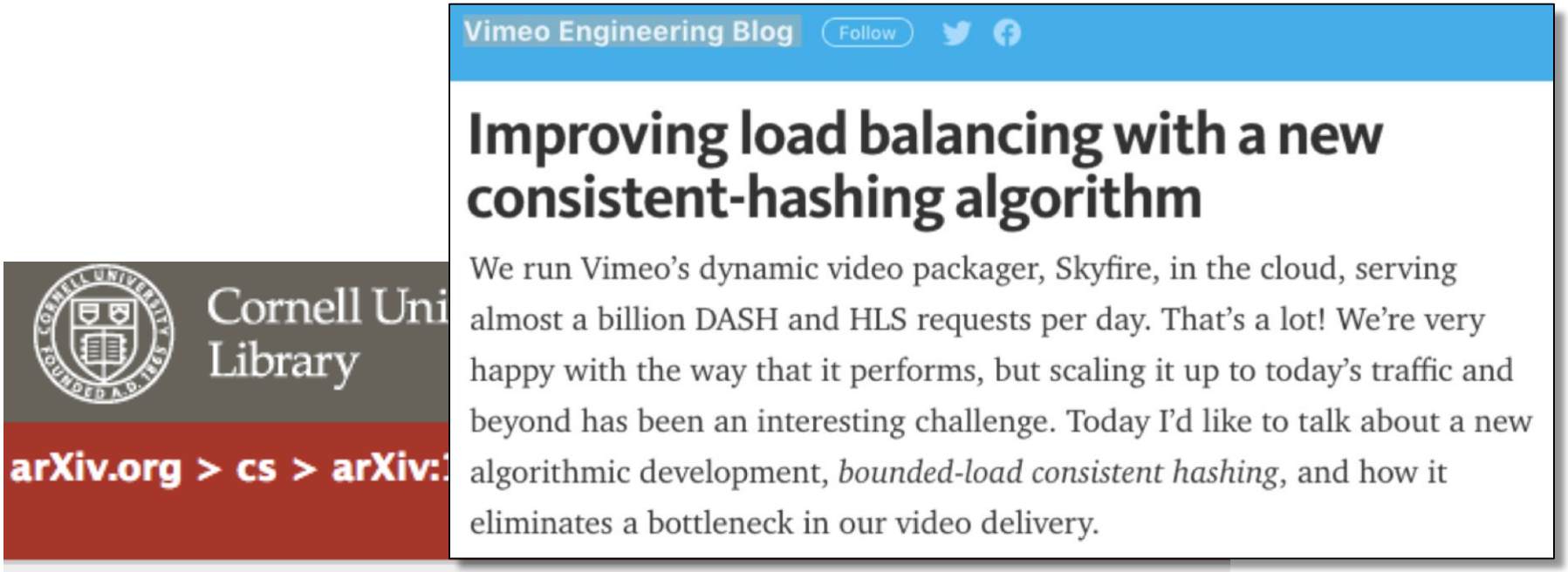
Three features that distinguish Chord from many other peer-to-peer lookup protocols are its simplicity, provable correctness, and

Vimeo's bandwidth bottleneck

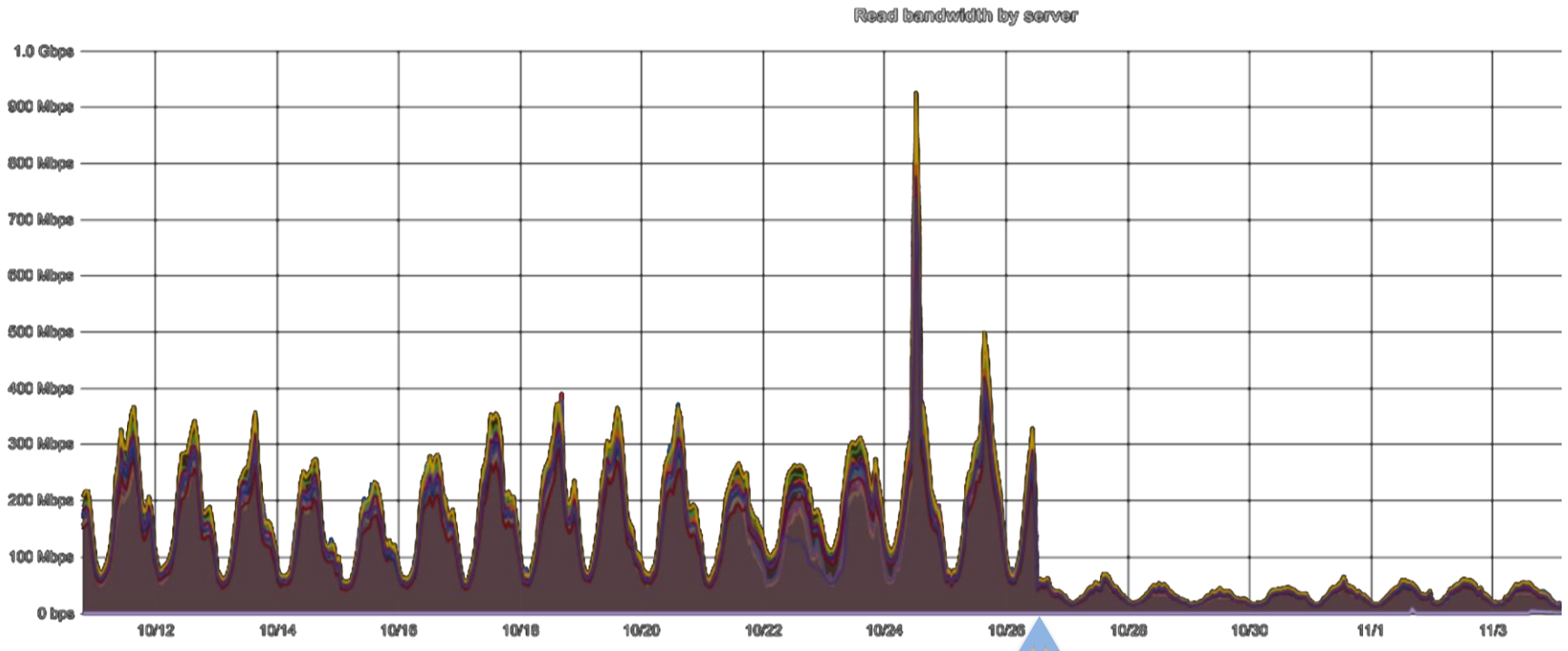


Issue: High bandwidth requirement…

From algorithm theory to industrial reality



Eliminating the bandwidth bottleneck



Old algorithm

New algorithm

switch

One algorithm, many applications

- Our algorithm has no details specific to video streaming. Works for *any* dynamic allocation system in the world – e.g. Google's cloud.

- Mathematical analysis based on properties of degree-4 polynomials with random coefficients – the theory of which was originally developed with other applications in mind.

Lemma 10. *The expected number of balls hashing directly to any expected number of balls forwarded into q from its predecessor q^- is not active, and its active successor q^+ is given an extra capacity of o bins starting from q^+ is $O((\log c)/c^2)$.*

Proof. For the first statement, we note that the expected number of ba n/r for any $0 \leq i \leq r$. These are not added to q if some bin hash to $[h$ event because balls and bins hash independently. The expected numb is $\mu = i(n-1)/r$. For $i \geq r/(n-1)$, we have $\mu \geq 1$, and then, by in $[h(q)-i, h(q))$ is $O((\mu+\mu^2)/(\mu-0)^4) = O(1/\mu^2) = O((r/($ hashing directly to q is thus bounded by

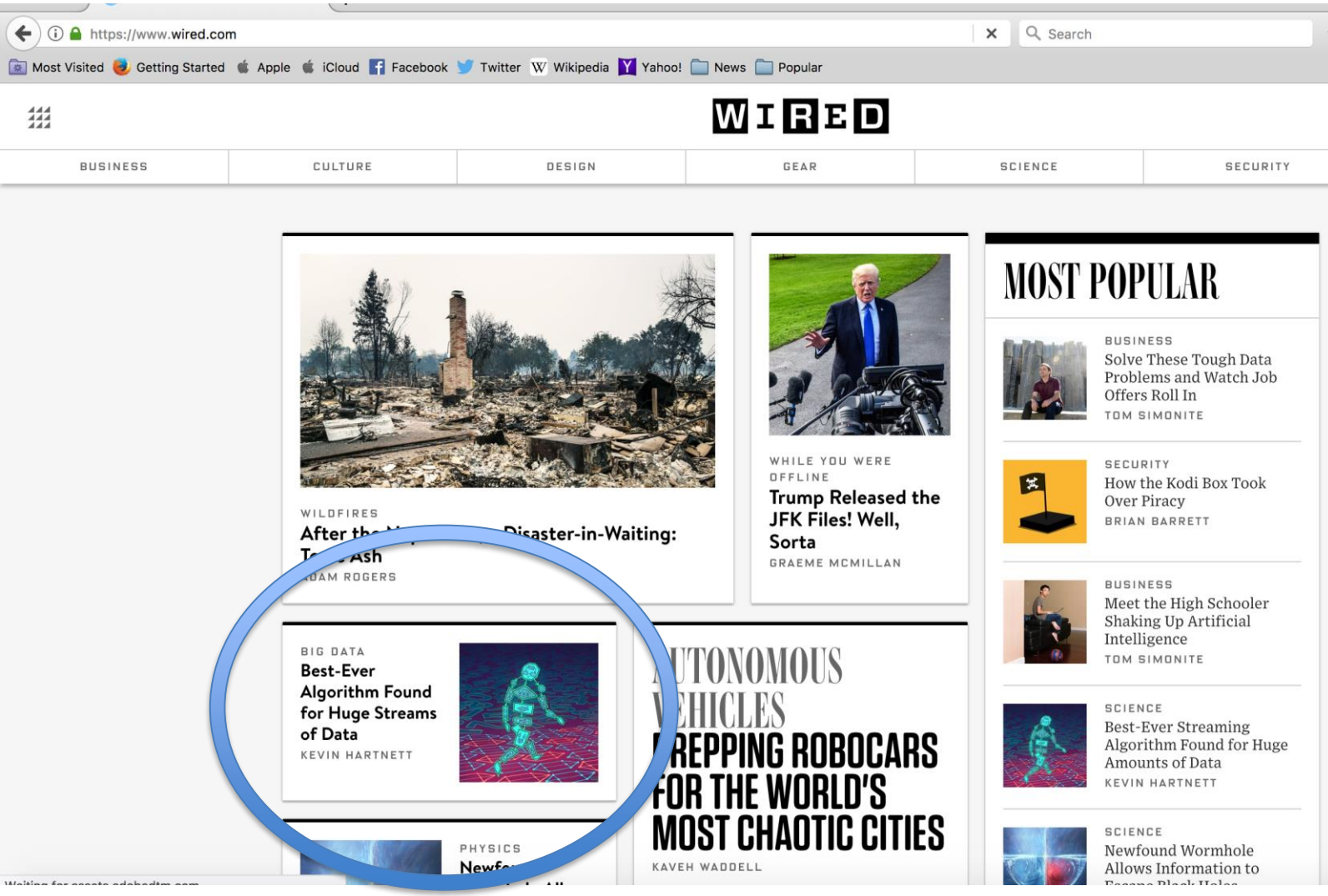
$$n/r \cdot \left(\lfloor r/(n-1) \rfloor + \sum_{i=\lfloor r/(n-1) \rfloor + 1}^{\infty} (r/(ni))^2 \right.$$

We also have to consider the probability that the preceding bin q^- fo we would need q^- to be filled even if we increased its capacity by 1 least 2. This is bounded by the probability of having an interval $I \ni$ bins including one with capacity at least 2. This is what we analyze $\Pr[d \geq 1] \leq \mathbf{E}[d] = O((\log c/c^2)$. By the capacity constraint, the forwarded to and end in q is $2cm/n$, so the expected number is

$$O((\log c/c^2)2cm/n = O((m/n)(\log$$

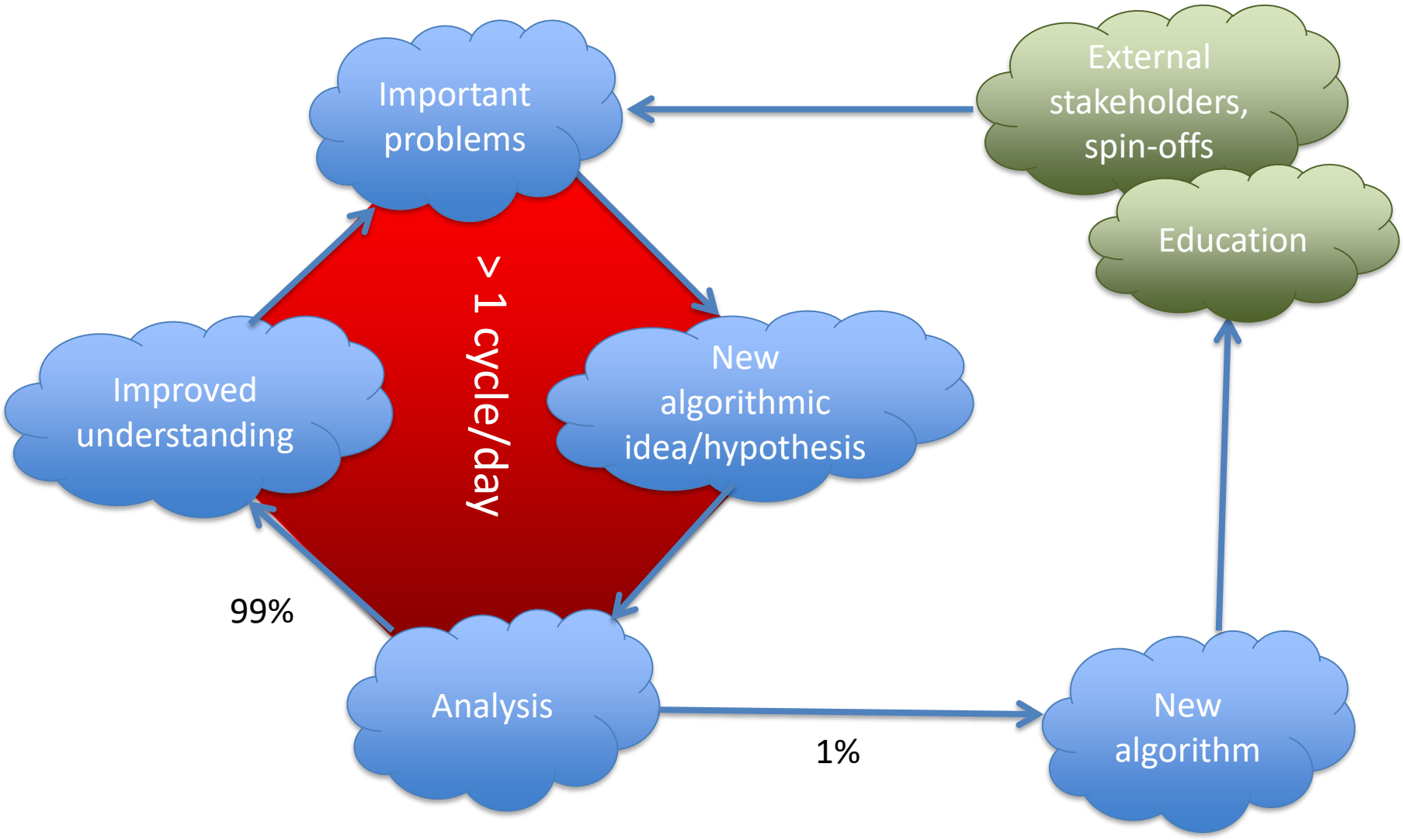
Next we ask for the expected number d of full bins starting from the bin q, when q^+ is given an extra capacity of one. Again this implies th the analysis from the proof of Lemma 9 implies that $\mathbf{E}[d] = O((\log c$

Press on recent work with
K.G. Larsen (Århus), J. Nelson (Harvard), Nguen (Northeastern)



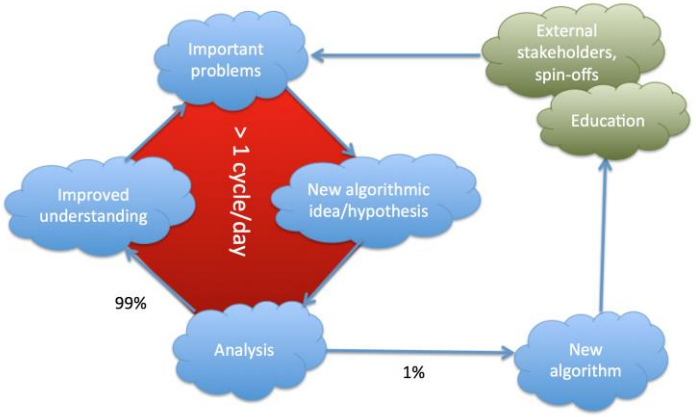
Nature of algorithmic research

- All research can be thought of as feedback loop,
  - Generating hypothesis/idea.
  - Experiment testing hypothesis, with outcome affecting next hypothesis.
- In experimental fields, experiments may be expensive and take years, so they better be carefully thought out.
- In algorithms we use mathematical thought experiments that may take a day or less…
  This changes everything!


Important problems
External stakeholders, spin-offs
Education
< 1 cycle/day
Improved understanding
New algorithmic idea/hypothesis
99%
Analysis
1%
New algorithm

Ingredients of successful algorithmic research

- Talent.
- Deep understanding and intuition.
- Meeting of great minds in a creative environment.
- Inspiring leadership, research activities.
- Courage to do unpredictable research.



Core team


Thorup


Husfeldt


Pagh


Alstrup

Full professors with proven track record:

- **Academic impact** (H-indexes 54, 16, 26, 21)
- **Visibility** (keynotes ICDT ‘15, CPM ‘15, ESA ‘15, ECML ‘16, HALG ’16, ICALP ’17,…; editorial board/PC memberships of leading journals/conf.; “digitale vismandsråd” member…; appeared in BBC, ZDF, DR)
- **Managing large research projects** (DFF Adv. Grant 2013-18, ERC Consolidator 2014-19, Inno+ Big Data 2016-20)
- **Industrial impact** (AT&T Fellow Honor; Science Innovation Prize; founding Octoshape w/100M users,…)
- **Nurturing talent** (our PhD students received Presburger award, Nerode Prize, EliteForsk stipends)

Unique strength to attract the best talent

Vision

Ideal research environment, attractive to the best researchers in the world.

Regarded as the leading center for algorithms in Europe

World-wide impact on science and industry

High impact on education

Effect multiplied by synergy with Danish algorithms community and industry

Flying start from DFF center Efficient Algorithms and Data Structures (EADS)

- Leadership in Algorithms
- ACM-SIAM Symposium on Discrete Algorithm (SODA) top international venue.
- 10 out of 180 papers accepted worldwide for SODA'18 from DIKU.
- I am 1 out of 4 invited plenary speakers.

PhD students

- 4 PhD students finished this year.
- Eva Rotenberg become Assistant Professor at DTU.
- Mikkel Abrahamsen joined 3Shape researching hearing aids.
- Søren Dahlgaard and Mathias Bæk Knudsen and doing machine learning start-up SupWiz.

Summary

- ICT is a vehicle for societal progress
- Designing efficient algorithms for fundamental problems is key to keeping up progress
- Algorithms research is best performed in a vibrant center where the best minds interact
- Unique strength of PI and core team can attract the best PhD students, post-docs, and visitors
- BARC aims to become the leading hub for algorithms research in Europe